

Nucleotide Sequence of a Human Cytomegalovirus DNA Fragment Encoding a 67-Kilodalton Phosphorylated Viral Protein

MICHELLE GALL DAVIS^{1*} AND ENG-SHANG HUANG^{1,2}*Lineberger Cancer Research Center¹ and Department of Medicine and Department of Microbiology and Immunology,² University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514*

Received 11 February 1985/Accepted 10 June 1985

Previously, we demonstrated that a DNA fragment from the L-unique region at 0.37 to 0.39 map units of human cytomegalovirus DNA codes for a 67-kilodalton phosphorylated tegument protein. (Davis et al., J. Virol. 52:129-135, 1984). We have determined the nucleotide sequence of this DNA, transcription initiation and termination, and putative promoter region. We also compared the properties of the protein product predicted from the DNA sequence with the properties of the protein isolated by immunoabsorbent column chromatography. This viral structural protein may have important functions, such as protein kinase activity, DNA binding, and possible transcriptional activation of immediate-early genes.

Human cytomegalovirus (HCMV) is a herpesvirus with cascading transcription, as described by several investigators (3, 11). One of the strongest regions of transcription late in infection is located in the center of the L-unique region (2). A DNA fragment which includes this transcription region was transferred to simian virus 40-transformed monkey kidney cells (COS-1) and a 67-kilodalton (67K) cytomegalovirus protein was detected in extracts of these cells by reaction of blotted proteins with virus-specific monoclonal antibody (2). The 67-K protein is a phosphorylated protein found in isolated virions. It probably corresponds to the 66K phosphorylated tegument protein described by Gibson (4) in the simian cytomegalovirus system.

The 67-K protein has been isolated by immunoabsorbent column chromatography (E.-C. Mar and E.-S. Huang, manuscript in preparation). It binds to DNA and phosphorylates proteins, an activity that was previously discovered in isolated virions (7) and in infected cells (9). In this communication, we present the nucleotide sequence of the coding region and mapping of the RNA transcript.

MATERIALS AND METHODS

Recombinant plasmids. Plasmid pHD7 (pBR322-HCMV *EcoRI*-G) and recombinant M13-713 were previously described (2). The 2.9-kilobase (kb) *Bgl*II fragment from plasmid pHD7 was inserted at the *Bam*HI site of the replicative-form DNA of M13mp7 to generate recombinant M13-713. The *Bgl*II-digested DNA of plasmid pHD7 and *EcoRI*-digested, replicative-form DNA of M13-713 were used for hybridizations and sequence analysis.

Sequence analysis. DNA was labeled for nucleotide sequence analysis by incorporation of α -³²P-labeled deoxynucleotide monophosphate at the 3' termini with large-fragment polymerase (Bethesda Research Laboratories) or by incorporation of [γ -³²P]ATP at the 5' termini with polynucleotide kinase. The chemical degradation method of Maxam and Gilbert (8) was used. A summary of labeled fragments is illustrated in Fig. 1. Enzymes were obtained from Bethesda Research Laboratories and digestions were carried out as recommended by the supplier.

RNA mapping. End-labeled DNA fragments prepared for sequencing were also used as primers for cDNA synthesis.

DNA primers were hybridized to RNA isolated from WI-38 cells infected with the Towne strain of HCMV and extended by reverse transcriptase to determine the location of the 5' end of the RNA transcript. RNA was isolated from infected cells by lysis with 7 M guanidine hydrochloride and Sarkosyl followed by centrifugation through 6 M CsCl at 34,000 rpm for 12 h in an SW41 rotor. RNA pellets were extracted with phenol and chloroform. Approximately 100 μ g of total RNA was mixed with 3' labeled DNA fragments. The mixture (25 μ l) was heated at 95°C for 5 min and slowly cooled to 40°C. In a final volume of 100 μ l, the RNA-DNA hybrids were incubated with 16 mM KCl-50 mM Tris hydrochloride (pH 8.3)-10 mM MgCl₂-0.1 mM four deoxynucleoside triphosphates-10 mM β -mercaptoethanol-100 μ g of bovine serum albumin per ml-3 μ l of reverse transcriptase (Life Sciences, Inc.). After incubation at 37°C for 3 h, samples were extracted with phenol. RNA was degraded by addition of 0.1 M NaOH and incubation for 1 h at room temperature, followed by addition of 0.15 M NaH₂PO₄. Nucleic acids were repeatedly ethanol precipitated to remove salts, suspended in 80% formamide, denatured by boiling, and loaded onto 4% acrylamide gels containing 7 M urea. cDNA was detected by autoradiography.

Total cellular RNA was also isolated from infected cells labeled with ³²P_i as described previously (2).

Nucleic acid hybridization. Nick translations and hybridizations were performed as described previously (2).

RESULTS

Nucleotide sequence of the 67K gene region. Enzymes *Hae*III, *Hinf*I, *Dde*I, *Alu*I, and *Taq*I were used to fragment the isolated 2.9-kb *Eco*RI fragment of M13-713 replicative-form DNA. These fragments were end labeled and chemically degraded. The sequence listed in Fig. 2 begins at the first determined nucleotide, 165 base pairs (bp) upstream of the leftward *Bgl*II site of Fig. 1, at 0.37 map unit. This sequence and the nucleotides 106 bp downstream of the rightward *Bgl*II site at 0.39 map unit were determined by labeling *Bgl*II fragments of plasmid pHD7 DNA. The orientation of these *Bgl*II fragments on the HCMV map was determined by hybridization to pHD7 DNA immobilized on nitrocellulose filters. The entire sequence, 3,072 bp, is listed in Fig. 2.

RNA mapping. A 239-bp fragment from bases 717 to 956

* Corresponding author.

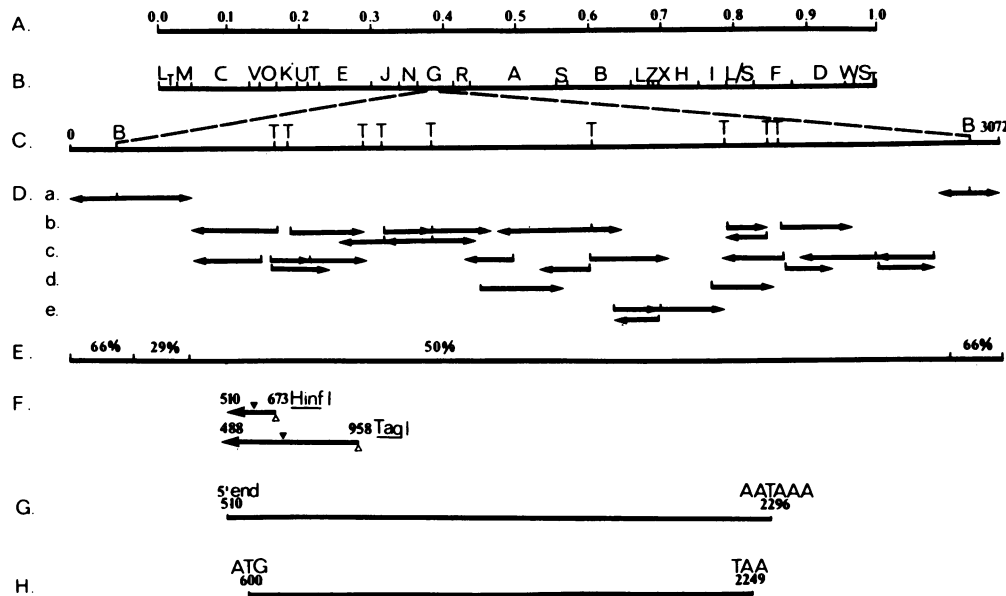


FIG. 1. Physical map of the HCMV DNA molecule. (A) Fractional length of the HCMV DNA molecule. (B) Recognition sites for *EcoRI* endonuclease on DNA isolated from HCMV Towne strain, passage 36, previously published (1). (C) Expanded region of fragment *EcoRI*-G, indicating the extent of the nucleotide sequence. 0, First nucleotide determined; 3072, last nucleotide determined. T, Recognition sites for *TaqI* endonuclease. B, Recognition sites for *BglII* endonuclease. (D) Labeled fragments (dots) and nucleotide sequence determinations (arrows) used to compile the sequence. Fragments labeled at: (a) *BglII* sites; (b) *TaqI* sites; (c) *HinfI* sites; (d) *DdeI* sites; (e) *HaeIII* sites. (E) G+C content of the DNA. (F) Fragments labeled at the 3' end (▼) and extended by reverse transcriptase (arrows): Δ, nucleotide corresponding to the 5' end of the priming fragment; ◀, nucleotide corresponding to the 3' end of the cDNA product. (G) Boundaries of the RNA transcript. (H) Boundaries of the protein translation region.

generated by *TaqI* digestion of M13-713 DNA was labeled by incorporation of [α - 32 P]dCMP at the 3' ends. After extension of this primer with reverse transcriptase, a cDNA product 470 bases long was detected (Fig. 3). This observation was used to estimate the location of the RNA initiation at base 488 (5' *TaqI* site at 958-cDNA product 470). A 62-bp fragment from bases 608 to 670 generated by *HinfI* digestion of M13-713 DNA was labeled by incorporation by [α - 32 P]dAMP. After extension of this primer with reverse transcriptase, a cDNA product 164 bases long was detected. This product was compared with *HinfI*-digested pBR322 and a partially degraded DNA fragment 230 bases long. The 3' end of this cDNA is at base 510, 164 bases from the 5' *HinfI* site at base 673, corresponding to the 5' end of the RNA. See Fig. 1 for a schematic illustration of this experiment.

The 3' end of the RNA is probably near base 2295, at sequence AATAAA. This is indicated by the hybridization of RNA to the fragmented DNA of plasmid pHD7 (Fig. 4). Digestion of DNA of plasmid pHD7 with enzyme *TaqI* generates an 1,150-bp fragment which includes the sequence between bases 0 and 668 (the 5' *TaqI* site is not included on the sequence), a doublet at 510 bp which includes sequences between bases 1025 and 2087, and a 760-bp fragment which includes sequences between bases 2304 and 3072 (the 3' *TaqI* site is not indicated on the sequence, although it must be close to the last determined nucleotide). The 1,150-bp fragment and the doublet at 510 bp hybridize with RNA isolated from infected cells, but the 760-bp fragment does not (Fig. 4). Thus, the RNA termination is 5' to base 2304, probably at base 2296, sequence AATAAA.

Protein coding. The initiation codon AUG occurs at base 548, again in phase at base 566, and the terminator TAG at base 605. We have no information about possible *in vivo* synthesis of this 19-amino acid oligopeptide.

The codon AUG also occurs at base 600 and the coding region continues for 549 amino acids to base 2246. The composition of the protein encoded by this sequence is very basic: 20% lysine, arginine, and histidine, 5.5% aspartic and glutamic acids, 36% hydrophobic residues, and 38% other residues. The 549-amino acid protein would be 66 kilodaltons, close to the size predicted by gel electrophoresis of the protein immunoprecipitated by monoclonal antibodies specific for this gene product. The basic composition of the protein is consistent with DNA binding activity.

DISCUSSION

HCMV DNA was mapped by electron microscopy of partially denatured DNA molecules (5), illustrating that some low guanine-plus-cytosine (G+C) stretches are easily found and other regions of high G+C composition are also present. The first 250 bp and the final 250 bp of this 3,072-bp stretch listed in Fig. 2 contain 66% G+C bases, but the sequence 250 to 500 bp, just upstream from the 5' end of the RNA, contains 29% G+C bases, and the sequence between 500 and 2840 bp, including the entire coding region, contains 50% G+C bases (Fig. 1E).

The initiation of RNA transcription at base 510 in the sequence listed in Fig. 2 was determined by primer extension of DNA fragments which end near the 5' end of the RNA. Total RNA was used as a template; this RNA contains a single 1.9-kb species which hybridizes with the primer fragments (2). Since the only radioactive label added was the priming fragment, the only labeled product was the cDNA resulting from extension of the primer. Indeed, the major product in two separate experiments was a single band with few intermediate or obviously degraded products (Fig. 3).

The sequences often observed 5' to the initiation of RNA transcription in eucaryotes are observed at typical locations

1:GCGG GCGC GCGA TGGT ATGT GCGC GCGA AGCA GCGC GCGC TGCG TGCG AGCG AAGC
 71:GCGG GCGA TGGC GCGC GCGC GCGC GCGC GCGC GCGC GCGC GCGC GCGC GCGC
 141:AGCG TGGT AGCG AAGT GCGC TCGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 211:GCGG TCGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 281:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 351:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 421:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 491:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 561:GCGG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 Met Gys Arg Ile Arg Thr Phe Arg Leu Arg Arg Arg Val Ser Gly Arg Ser Phe
 600:ATG TGT AGG ATT CCG ACT TTC COT TGC AGA AGA AGA AGA CCG GTC TCT GGA CCG TCT TTT
 Gys Arg Ser Arg Ile Asp Pro Gly Tyr Val Arg Glu Arg Pro Tyr Ile Gly Gly Arg Ser
 660:TGT CCG TCT GCA ATC GAC CCG GGA TAC GGA AGA GAG CCG CCG TAC ATC CCG CCG TCG
 Arg Pro Thr Thr Phe His Leu Thr Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg
 720:AGA CCG AGC AGC TTC CAG CAG AGC AGA AGA AGA AGA AGA AGA AGA AGA AGA AGA
 Pro Ser Phe Ala Arg Ser Pro Val Leu Glu Pro Ser Ser Pro Pro Glu Ala Leu Ser Val
 780:CGA TCG TTT CCG CCG CCG GTC CTC GAG CGA TCA TCA CCA CCG CCG CCG CCG CCG
 Pro Ser Leu Ser Ser Glu Lys Lys Thr Ala Ser Pro Thr Gys Val Lys His Leu Ser
 840:CGA TCG TGC TCA TCC GAA AAA AAA ACT GCG TCA CCG ACC ACC ACC ACC ACC ACC ACC
 Gly Gly Ala Val Arg Arg Arg Val Val Val Pro Arg Thr Lys Lys Lys Lys Lys Lys Lys
 900:CGA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 His His Thr Leu Met Ser Thr Thr Gys Arg Tyr Trp Ser Ser Ser Ile Val Leu Thr Glu
 960:CGC CAC ACC CTC ATG AGC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC
 1020:CAT CTC GAC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Gly Gly Thr Leu Ser Gly Ser Ile Ile Ser Ile Val Ala Lys Arg Arg Arg Arg Arg Arg
 1080:CGC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Glu Pro His His Leu Phe Phe Met Leu Leu Tyr Tyr Lys Arg Leu Met Ser Thr Thr Leu
 1140:CGA CCA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Gly Ile Ser His Leu Val Lys His Ala Arg Phe Ser Ser Leu Glu Gly Thr Leu Gly Ser
 1200:CGC ATC ACC CAG CTC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Leu Lys Gly Thr Val Tyr Arg Pro Lys Glu Lys Thr His Thr Gys Arg Arg Ser Thr Thr
 1260:CTG AAA CCG ACC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Ser Arg Asp Thr Lys Thr Glu Ser Tyr Arg Ala Leu Leu Met Glu Arg Thr Trp Arg Lys
 1320:TCG AAT GAC AGC AGA AGA CCA TCG TTT AAT CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Lys Thr Ile Ala Ala Arg Ala Lys Pro Lys Pro Ser Gly Glu Ala Gly Ala Ser Phe Pro
 1380:AAA ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC
 Thr Glu Ser Leu Gly Ser Leu Ser Ser Phe Phe Pro Arg Ser Leu Lys Arg His Arg
 1440:ACT CAG TCT CTC GCG TCG TTT ACC TCT TTT TTT CCG CAG TCT TCT CCG AGG AAT CAG CCG
 His Ser Pro Met Thr Arg Arg Glu Leu Leu Glu His Lys Arg Ser Glu Val Glu Thr
 1500:CGC ACC CCG ATC AGC CCA AAT CCA TGC CTA GAA CAA CCG CCG CCG CCG CCG CCG CCG
 Arg Lys Leu Arg Arg Ala Ile Thr Leu Tyr Gys Val Tyr Ile Arg Arg Gys Glu Thr Val
 1560:AAC AAA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Gly Lys Lys Arg Arg Gys Met Met His Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg
 1620:CGT AAA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Pro Arg Ala Arg Thr Glu Thr His Arg Phe Pro Ser Pro His Lys Glu Met Arg Pro
 1680:CGC AAT GCA AAT CCA CCA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Pro Thr Arg Glu Ala Asp Ser Gys Ser Ser Tyr Asp Gly Ala Leu Ala Arg Arg Ala Ser
 1740:CGC AAT GCA AAT CCA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Ser Pro Leu Arg Glu Gly Thr Ser Gly Ser Ser Ala His Arg Pro Lys Glu Glu Lys Leu
 1800:TCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Leu His Gys His Glu Ser Ile Ser Gys Ile Leu Phe Pro Pro Arg Ser Ser Pro Glu His
 1860:CGA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Glu Phe Ala Leu Leu Lys Lys Val Val Val Ala Thr Ala Ala Thr Gly Arg Ser Ser
 1920:CGA TTC CCA CTA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Ser Pro Arg Pro Glu Thr Leu Tyr Met Gys Arg Lys Lys Val Arg Thr Gys Tyr Gys Pro
 1980:TCT CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Met Arg Leu Val His Arg Tyr Met Phe Glu Arg Leu Leu Ile His Leu Arg Ala His Arg
 2040:ATG CCA CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG
 Leu Gly Val Ser Ser Ser Ser Arg Trp Pro Thr Phe Ser Thr Thr Val Val Arg Pro Gys
 2100:TTA GCA GGT AGT TCC TCC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC
 Thr Thr Thr Pro Ile Lys Thr Thr Ser Ser Ser Ser Pro Arg Pro Arg Arg Arg Arg
 2160:ACT ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC
 Ser Pro Pro Ile Leu Gly Met Gys Phe Ser
 2220:TCG CCG CCA ATC TGC GCG ATG TGT TTT TTT
 2281:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2341:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2401:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2461:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2521:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2581:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2641:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2701:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2761:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2821:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2881:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 2941:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3001:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3061:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3121:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3181:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3241:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3301:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3361:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3421:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3481:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3541:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3601:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3661:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3721:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3781:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3841:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3901:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 3961:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4021:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4081:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4141:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4201:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4261:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4321:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4381:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4441:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4501:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4561:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4621:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4681:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4741:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4801:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4861:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4921:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 4981:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5041:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5101:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5161:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5221:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5281:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5341:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5401:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5461:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5521:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5581:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5641:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5701:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5761:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5821:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5881:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 5941:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6001:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6061:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6121:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6181:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6241:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6301:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6361:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6421:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6481:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6541:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6601:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6661:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6721:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6781:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6841:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6901:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 6961:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7021:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7081:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7141:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7201:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7261:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7321:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7381:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7441:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7501:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7561:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7621:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7681:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7741:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7801:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7861:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7921:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 7981:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8041:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8101:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8161:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8221:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8281:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8341:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8401:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8461:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8521:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8581:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8641:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8701:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8761:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8821:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8881:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 8941:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9001:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9061:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9121:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9181:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9241:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9301:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9361:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9421:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9481:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9541:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9601:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9661:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9721:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9781:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9841:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9901:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 9961:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG
 10001:AGCG TGGC AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG AGCG

FIG. 2. Nucleotide sequence of HCMV DNA, 0.37 to 0.39 map units. The sequence of the anticoding strand is written 5' to 3'. The interpreted amino acid sequence is indicated in the open reading frame from base 600 to 2249. Underlined bases are the 5' end of the RNA at base 510 and the polyadenylation site at base 2295.

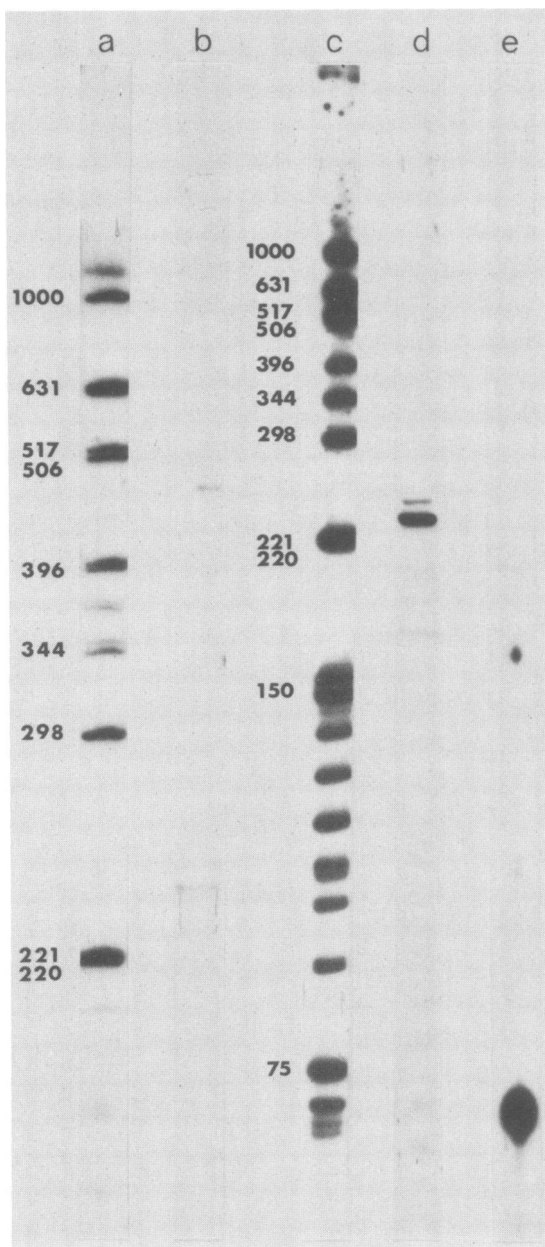


FIG. 3. Primer extension of 3' end-labeled DNA fragments. X-ray exposures of 4% acrylamide gels containing fragments of pBR322 plasmid DNA digested with *Eco*RI and *Hinf*I, labeled with [α - 32 P]dAMP and large-fragment polymerase (lanes a and c), cDNA primed with a 239-bp fragment (lane b), a partially degraded DNA fragment (lane d), and cDNA primed with a 70-bp fragment (lane e). Additional fragments in lane c correspond to M13mp7 DNA. The fragment length in bases for lane a is indicated to the left, and the fragment length in bases for lane c is indicated to the right.

5' to the initiation of transcription at base 510 of this sequence. The sequence TATATATATA occurs at base 24 (486 bases upstream), ATATATG occurs at base 283 (228 bases upstream), TATATAG occurs at base 385 (126 bases upstream), CAAT occurs at base 452 (58 bases upstream), and ATATATATG occurs at base 475 (45 bases upstream). Some alignment of related sequences can be drawn which includes pyrimidine tracts and alternating AT sequences; examples are illustrated in Fig. 5. Many hairpin loop struc-

tures can be drawn which involve these sequences; the significance of such structures is unknown. The direct repeats reported by Stinski et al. (10) were not found, and no long direct repeats were detected by computer analysis. This DNA region is only transcribed in infected cells after the initiation of viral DNA synthesis, and WI-38 cells treated with recombinant plasmids including this coding region do not produce the 67K protein. We presume that an α or β viral gene product is required to stimulate expression of this viral gene.

The termination of the RNA transcript near base 2295 is indicated by the sequence AATAAA. The RNA isolated from infected cells does not hybridize with the 760-bp DNA fragment that includes the sequence from base 2304 to base 3072, generated by *Taq*I digestion of pH7 DNA. Thus, the

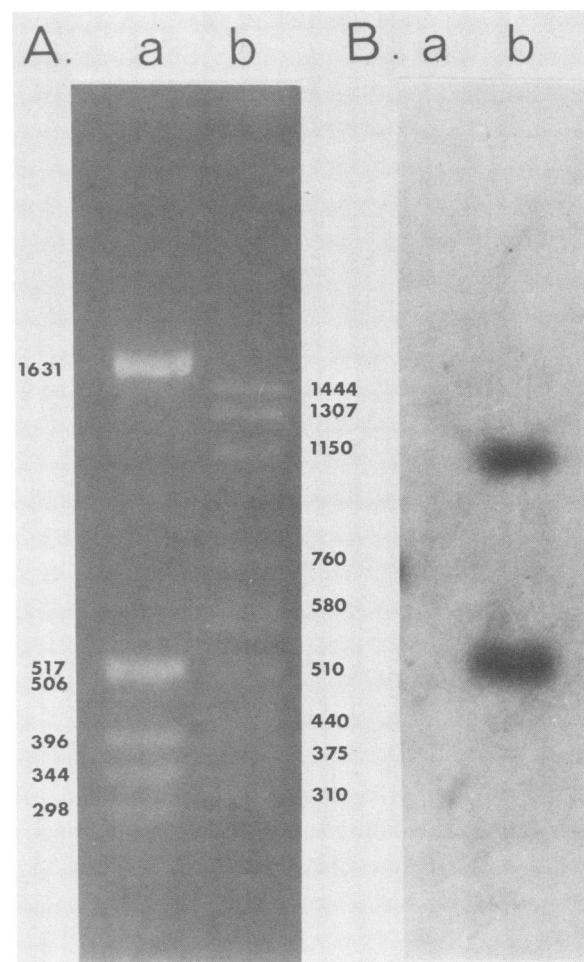


FIG. 4. Hybridization of HCMV RNA to DNA of plasmid pH7. A 2% agarose gel (A) containing DNA fragments of plasmid pBR322 digested with *Hinf*I (lane a) and plasmid pH7 digested with *Taq*I (lane b) was blotted onto nitrocellulose paper which was hybridized with 32 P-labeled RNA isolated from WI-38 cells infected with HCMV. Fragment lengths in lane a are 1,631, 517, 506 (doublet), 396, 344, and 298 bp. Smaller fragments are smeared on this gel. Fragments corresponding to pBR322 sequences in lane b are 1,444, 1,307, 475, 368, and 312 bp. Many of the stained bands smaller than 1 kb contain several comigrating species. (B) An X-ray exposure of the nitrocellulose paper after hybridization. Hybridizing fragments correspond to a 1,150-bp fragment and a 510-bp doublet in lane b.

First base:	Sequence:	
69	CCCCCCCCCATATTTACGGC	
248	GTCCCGCTATATTTCTATT	
6	GTGCCCCCTATATATATATT	
466	TATCCCTCCATATATATGAT	
377	TACGCTCTTATATAGTTCTC	
	5 - 9 / 4 - 12	
	pyrimidine / AT	
	mostly C /	
266	TATTTTTTTTCTACAAAATATATGATGAAC	
316	TCGTCTATTTTTA	ATATGCTCTAC
513	CCTTTTTTTTCT	CATCTCCTGTAT
	12 / 11 - 12	
	pyrimidine / related bases	
	mostly T /	

FIG. 5. Alignment of repeated sequences occurring 5' to the initiation of transcription.

RNA must terminate 5' to base 2304. There are no apparent splices in the sequence, and our RNA-DNA heteroduplexes observed by electron microscopy do not contain any "R loops," indicating no large spliced regions. The mRNA from infected cells which hybridizes with this DNA is 1.9 kb (2), which would include 1,785 bases from bases 510 to 2295, plus 100 A bases at the 3' end.

The protein-coding region indicated by the sequence is sufficient for the protein that we have determined is coded by this DNA fragment. We have no amino acid sequence data on this protein product. The monoclonal antibody which identifies the 67K protein in extracts of monkey kidney cells carrying the DNA was used to isolate the 67K protein by immunoabsorbent column chromatography. The protein kinase activity eluted with the 67K protein (Mar and Huang, in preparation). Michelson et al. (9) also reported that the protein kinase isolated from HCMV-infected cells was 68 kilodaltons.

We do not detect transcription of this coding region in HCMV-infected cells incubated in cycloheximide, although the 67K protein is observed in infected cells after uptake of the virus. We believe the protein is taken into the cell along with the viral DNA. It is a basic protein with protein kinase activity, it binds to DNA, and it may be the structural component which functions as an initiator or key enzyme to phosphorylate an initiator of immediate-early gene transcription (10). A similar function has been suggested for the major tegument protein of herpes simplex virus (1).

Protein kinase activity has been measured in isolated HCMV virions (7) and herpes simplex virus virions (6). In HCMV, the protein kinase is apparently a tegument protein (2). The ratio of virus particles/PFU is estimated to be more than 500:1, and those defective virus particles are packed with tegument proteins: thus, thousands of molecules of protein kinase are transferred to each infected cell.

HCMV-infected cells carry on a significant level of host cell protein synthesis, while herpes simplex virus-infected cells do not. Perhaps an enzyme function, such as the viral protein kinase, carried into the infected cells will alter the activity of the host cell machinery. A comparison of the protein kinase functions coded by these viruses may provide some reason for this difference.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants CA21773 and CA19014 from the National Cancer Institute and grant AI15036 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. Campbell, M. E. M., J. W. Palfreyman, and C. M. Preston. 1984. Identification of herpes simplex virus DNA sequences which encode a *trans*-acting polypeptide responsible for stimulation of immediate-early transcription. *J. Mol. Biol.* **180**:1-19.
2. Davis, M. G., E.-C. Mar, Y.-M. Wu, and E.-S. Huang. 1984. Mapping and expression of a human cytomegalovirus major viral protein. *J. Virol.* **52**:129-135.
3. Demarchi, J. M. 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early, and late RNAs. *Virology* **114**:23-38.
4. Gibson, W. 1981. Structural and nonstructural proteins of strain Colburn cytomegalovirus. *Virology* **111**:516-537.
5. Kilpatrick, B. A., and E.-S. Huang. Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. *J. Virol.* **24**:261-276.
6. Lemaster, S., and B. Roizman. 1980. Herpes simplex virus phosphoproteins. II. Characterization of the virion protein kinase and of the polypeptides phosphorylated in the virion. *J. Virol.* **35**:798-811.
7. Mar, E.-C., P. C. Patel, and E.-S. Huang. 1981. Human cytomegalovirus-associated DNA polymerase and protein kinase activities. *J. Gen. Virol.* **57**:149-156.
8. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:49-560.
9. Michelson, S., M. Tardy-Panit, and O. Barzu. 1984. Properties of a human cytomegalovirus-induced protein kinase. *Virology* **134**:259-268.
10. Stinski, M. F., R. M. Stenberg, and W. F. Goins. 1984. Structure and function of the human cytomegalovirus immediate early genes, p. 399-421. *In* F. Rapp (ed.), *Herpesvirus*. Alan R. Liss, Inc., New York.
11. Wathen, M. W., and M. F. Stinski. 1982. Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early, and late times after infection. *J. Virol.* **41**:462-477.